



# 논문/리포트 요약문 생성 모델 구축 및 서비스 배포 Text Summarization Model

AI 16 심지은

# 목차

## I. 프로젝트 개요

## II. 프로젝트 구성

## III. 프로젝트 수행 절차 및 방법

## IV. 프로젝트 수행 결과

- a. 데이터 EDA
- b. Task Fit 모델 선정 및 학습
- c. 최종모델 선정
  - a. Score 개요
  - b. Rouge, RDASS
  - c. Rouge, RDASS - Candidates 비교
  - d. Human Evaluation
  - e. Human Evaluation – Candidates 비교
- d. 패키징 및 배포
- e. 최종 서비스

## V. 자체 평가 의견

## I. 프로젝트 개요

주제	논문/리포트 요약문 생성 모델 Text Summarization Model 제작 및 패키지 배포
의의	< 논문제공자 > 요약 자동화를 통한 사용자 편의성 향상 < 논문독자 > 논문 내용을 살피는 단계('검색화면'-'논문화면' 2단계) > '검색화면' 1단계를 줄임으로써 논문 검색시간 단축, 타사에 비한 검색 경쟁력 향상
목적	요약문을 수기 입력하지 않고 자동으로 생성하는 논문/리포트 자동 요약 서비스 제안
사용 예시	<div><div><div>"코로나19"검색결과 (11,277 건)</div><div><div><input type="checkbox"/> ⚡ 다운로드 <input type="checkbox"/> 내서재 담기</div><div>✓정확도순 이용순 최신순 20개씩 ▼</div></div><div><div><input type="checkbox"/> <b>코로나19</b> 시기 서울시 산업집적지 변화와 향후 대응전략</div><div><b>초록의 일부</b></div><div><div>코로나19발 경제위기, 팬데믹이 원인... 'K자형 경기침체·회복' 시나리오 제기코로나19발 경제위기는 기존 경제학 충격과 달리 글로벌 팬데믹 확산과 방역조치에 따른 수요와 공급의 동시 위축을 계기로 전 지구적으로 동시에 ...</div><div>김묵한, 조달호, 장윤희   서울연구원   서울연구원 정책과제연구보고서   2022.12   1-69(69pages) 연구보고서   이용수 8</div></div><div><div>📖</div><div>↓</div><div>🔖</div></div></div><div><div><div>"코로나19"검색결과 (11,277 건)</div><div><div><input type="checkbox"/> ⚡ 다운로드 <input type="checkbox"/> 내서재 담기</div><div>✓정확도순 이용순 최신순 20개씩 ▼</div></div><div><div><input type="checkbox"/> <b>코로나19</b> 시기 서울시 산업집적지 변화와 향후 대응전략</div><div><b>요약문</b></div><div><div>코로나19 사태 초반 경기 침체와 회복은 단기적 회복 형태로 여러 시나리오가 제안되고 검토되었지만, 최근 새로운 유형인 'K자형 경기 침체' 혹은 'K자형 회복(K-Shaped recovery)'이라는 개념이 코로나19 위기를 설명하는 시나리오로 등장했다.</div><div>김묵한, 조달호, 장윤희   서울연구원   서울연구원 정책과제연구보고서   2022.12   1-69(69pages) 연구보고서   이용수 8</div></div><div><div>📖</div><div>↓</div><div>🔖</div></div></div></div></div></div></div>

## II. 프로젝트 구성

담당업무

심지은

- 데이터 EDA, 정제 및 전처리
- 딥러닝 모델 선정 및 Fine Tuning
- 최종모델 선정을 위한 Evaluation Metric 구축 및 평가
- 패키징 및 배포를 위한 구조 설계

### III. 프로젝트 수행 절차 및 방법

#### EDA

DataSet 훑아보기

- 추출적 요약과 추상적 요약의 생성 방식 및 품질 비교
- Text 및 Label 선정

데이터 EDA

- 글자 수 확인
- 모델 요구 형태에 맞게 가공
- 데이터 3way-Split

#### Task Fit 모델 선정 및 학습

사전학습 모델 선정

- 한국어 텍스트 요약 Task에 적합한 KoBART 모델 사용

Fine Tuning

- 에포크 수 증가를 통해 성능 향상

모델 결과 도출

- HuggingFace Library의 Tokenizer 및 Method를 사용하여,
- 사전학습 모델 및 Fine Tuning 모델의 결과 도출

#### 최종 모델 선정

Evaluation Metric 구축 및 평가

- Rouge Score
- RDASS Score

Human Evaluation

정량지표(Rouge, RDASS)  
및 정성지표(Human)를 통해  
최종모델 선정

#### 패키징 및 배포

패키지 구조 구축

- 서비스 프로세스 구조화
- Fine-tuned Model 탑재 및 Utils 모듈 설계

Product Manual 작성

- 비전문가 대상으로 README 작성

GitHub Repo를 통한 배포

- Git-LFS를 이용하여 대용량 모델 탑재

## IV. 프로젝트 수행 결과 – EDA

DataSet	뉴스기사	사설	AIHUB / 요약문 및 레포트 생성 데이터 <ul style="list-style-type: none"><li>- 각 분류 별 데이터 200개 중 구어체로 쓰인 나레이션, 연설문, 회의록과, 목적이 다른 문학,역사문화재를 제외한</li><li>- 뉴스기사, 보도자료, 보고서, 사설, 간행물 총 5분류 1000개 자료</li><li>- 각 Data는 자료문서(Text)와 Abstractive생성요약문과 Extractive추출요약문, 2개의 Label로 구성되어 있음</li></ul>
	보도자료	간행물	
	역사문화재	연설문	
	보고서	문학	
	회의록	나레이션	
추출적 요약과  추상적 요약의  생성방식 및 품질비교	<b>추출적 요약 Extractive Summarization</b>  생성방식 문장 단위로 중요도를 scoring한 후 이를 기반으로 문서 내 중요문장을 선택/조합하는 비지도학습 모델 TextRank 등  뉴스기사, 소논문 등의 명확한 문장이 있는 요약 task에 적합함 그러나 문서 내 문장을 변형없이 사용하기에, 요약문장의 품질이 떨어져 사용자에게 어색함을 줄 수 있음		데이터 EDA
	<b>추상적 요약 Abstractive Summarization</b>  생성방식 원문을 기반으로 하되, 새로운 텍스트를 생성해내는 NLG natural language generation 지도학습 모델 Transformer/BERT 등  다양한 NLP Task에 활용되며, 각 Task마다 적합한 모델이 존재함 사전학습모델을 활용하여 FineTune 하여 성능을 향상시킬 수 있음		
Preprocessing <ul style="list-style-type: none"><li>- 모델 Input 요구 형태에 맞게, 전체 Text를 한 문장 형태로 가공</li></ul>			
데이터 3 way-Split <ul style="list-style-type: none"><li>- DataSet을 train(0.70), validation(0.15), test(0.15) 으로 분할하여 contamination 방지</li></ul>			
Tokenizer <ul style="list-style-type: none"><li>- 토크나이저 요구사항에 부합하는지(1024글자 이하), Test DataSet의 글자 수 확인</li></ul>			

Dataset의 생성요약문 Label을 활용하여, 추상적요약 사전학습모델을 FineTune하는 모델로 서비스 구축

## IV. 프로젝트 수행 결과 – Task Fit 모델 선정 및 학습

사전학습 모델 선정	<div>한국어 텍스트 요약 Task에 적합한 KoBART 모델 사용</div> <p><b>BART Bidirectional and Auto-Regressive Transformers</b></p> <ul style="list-style-type: none"><li>- 현 Text Summarization 분야의 SOTA 모델로, 입력 텍스트 일부에 노이즈를 추가 text Infilling 하여 이를 다시 원문으로 복구하는 autoencoder 형태로 학습하는 구조</li></ul> <p><b>KoBART</b></p> <ul style="list-style-type: none"><li>- Text Infilling 노이즈 함수를 사용하는 BART 모델에, 한국어 텍스트를 학습한 한국어 encoder-decoder 언어 모델</li><li>- 40GB 이상의 한국어 텍스트(한국어 위키 백과 이외, 뉴스, 책, 모두의 말뭉치 v1.0(대화, 뉴스, ...), 청와대 국민청원 등) 에 대해서 학습함</li></ul>
Fine Tuning	<div>자체 데이터셋을 추가로 학습하고 Epoch 수를 증대하여 Fine Tuning</div> <p>Merge Data</p> <ul style="list-style-type: none"><li>- 사전학습모델에 train, val Dataset을 추가로 학습시켜, Pytorch Framework를 통해 Fine Tune</li></ul> <p>Fine Tuning</p> <ul style="list-style-type: none"><li>- 에포크 수 증가를 통해 성능 향상</li><li>- Pytorch lightening Framework를 통해 각 에포크의 하이퍼파라미터 파일 hparams.yaml 과 학습된 모델의 가중치 체크포인트 .ckpt 생성</li><li>- 각 에포크마다의 가중치바이너리bin 및 구성요소json 을 얻어 각 모델 호출, 성능 평가</li></ul> <div>총 epoch 10번 Ep00 train_loss=1.380, val_loss=1.370 ... Ep03 train_loss=1.450, val_loss=1.340 ... Ep09 train_loss=1.060, val_loss=1.400 Ep10 train_loss=0.821, val_loss=1.430</div>
모델 결과 도출	<div>HuggingFace Library 의 Tokenizer 와 Fine-Tune Model 을 활용하여 총 10개 모델 및 사전학습모델의 요약결과 도출</div> <p>모델 결과 도출</p> <ul style="list-style-type: none"><li>- HuggingFace Library의 Method를 사용하여, FineTune Model('./model)과 KoBART Tokenizer ('gogamza/kobart-summarization ')을 호출하여 각 모델의 결과를 확인하고 성능 평가</li><li>- Fine-Tune하지 않은 pre-trained model과의 성능 비교</li></ul>

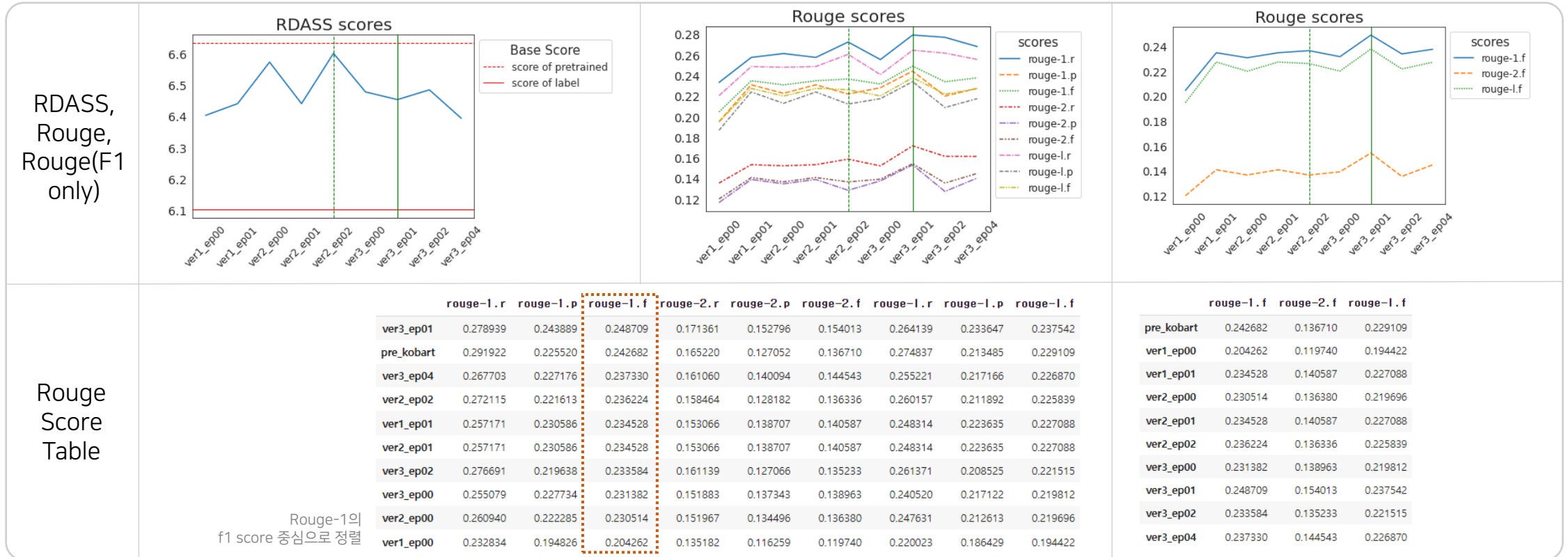
## IV. 프로젝트 수행 결과 – 최종모델 선정/Score 개요

텍스트 요약 Task의 Evaluation Metric	<p><b>Rouge Score</b></p> <ul style="list-style-type: none"><li>- 텍스트 요약 Task의 품질을 평가하는데 사용되는 일반적인 지표</li><li>- 기계생성 요약문과 정답문 사이의 n-그램(단어의 연속 시퀀스)의 중첩을 계산하여 유사성을 측정하며, n-gram에 따라 Rouge-1(유니그램), Rouge-2(바이그램), Rouge-L(최장공통시퀀스)으로 나뉨</li></ul> <p><b>RDASS Score</b></p> <ul style="list-style-type: none"><li>- 기계가 생성한 문장과 정답 문장의 유사도를 SBERT로 임베딩한 후 나오는 벡터 간 유사도로 측정하는 지표</li><li>- 단어 단위가 아닌 전체론적으로 평가하여, 교착어인 한국어 평가에 더 적합한 지표로 소개됨</li></ul> <p><b>Human Evaluation</b></p> <ul style="list-style-type: none"><li>- Rouge, RDASS 등 정량적 지표 외에도, 요약문의 표현적 다양성(diversity)을 위해 정성적 평가human evaluation 도 함께 수행함</li></ul>
Evaluation Metric 구축	<ul style="list-style-type: none"><li>- Rouge Score 패키지를 통해 Rouge Score 도출</li><li>- Sentence-Transformer의 SBERT를 불러와 문서(d), 정답문장(r), 모델생성요약문(p) 각각의 벡터 추출 후, <math>\langle Vp-Vr \rangle</math>과 <math>\langle Vp-Vd \rangle</math>의 코사인 유사도를 계산하여 두 유사도의 평균을 구하여 RDASS Score 도출</li></ul> <p>※ RDASS 논문에 따르면 SBERT를 미세조정fine tune하면 Score가 Human Evaluation에 더 가까워지나, 시간관계 상 미세조정없이 도출</p>



## IV. 프로젝트 수행 결과 – 최종모델 선정/Rouge, RDASS Score 결과

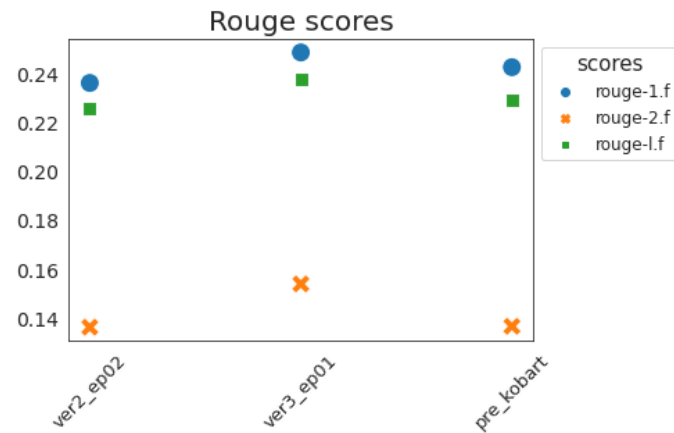
- 에포크 수를 증가시키면 일정 수준 성능이 올라갔으나, 에포크 수를 늘린다고 반드시 Score 향상으로 이어지진 않음
- RDASS score 중심으로 살폈을 때, Epoch 5(ver02\_ep02)의 성능이 가장 높았음
  - Fine-tune Model의 RDASS Score는 문서와 정답문Label의 유사도를 뛰어넘어 문서와의 유사성을 잘 추적하고 있다고 볼 수 있으나,
  - Pretrained Model의 RDASS Score에는 미치지 못함
- Rouge-1, 2, L의 f1-score 중심으로 살폈을 때, Epoch 7(ver03\_ep01)의 성능이 가장 높았음
  - Epoch 7의 Rouge-1의 F1 score는 Pre-trained Model의 Score보다 높았으나, 차이가 근소하여 통계학적으로 유의미하지 않음



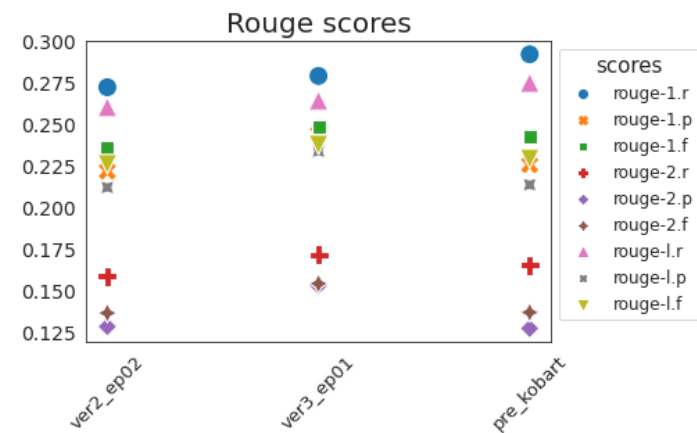
## IV. 프로젝트 수행 결과 – 최종모델 선정/Candidates 비교

- Rouge Score를 중심으로 Pretrained Model, Epoch 5(ver2\_ep02), Epoch 7(ver3\_ep01)의 학습 결과를 비교
  - Epoch 7 모델은 Pretrained Model에 비해 Rouge-1의 F1-score가 근소하게 높았고, 다른 Rouge 지표의 점수 결과도 Pretrained model에 비해 일정한(모여 있는) 양상을 보임

Pretrained, Epoch5, Epoch 7의 Rouge Score (F1 Score Only)



Pretrained, Epoch5, Epoch 7의 Rouge Score



# IV. 프로젝트 수행 결과 – 최종모델 선정/Human Evaluation

Sample Text	본 연구는 코로나19 이후 초등학생의 신체활동량과 신체적자기개념, 그리고 긍정심리자본의 관계를 탐색하고자 하였다. 구체적으로 운동참여가 긍정심리자본에 영향력을 행사하는 과정에서 신체적자기개념의 매개효과를 검증하고자 하였다. 이를 위해 초등학교 6학년에 재학 중인 97명의 학생들을 대상으로 주간운동빈도, 주간운동시간, 신체적자기개념, 긍정심리자본을 측정하는 설문조사를 실시하였고, 수집된 자료를 바탕으로 기술통계분석, 상관분석, 구조방정식 모형 검증을 실시하였다. 본 연구를 수행한 결과, 초등학생들의 운동참여수준(주간운동빈도, 주간운동시간)은 신체적자기개념 및 긍정심리자본의 모든 하위요인들과 정적 상관관계가 나타났다. 또한 신체적자기개념은 운동참여수준과 긍정심리자본의 관계를 부분적으로 매개하는 것으로 나타났다. 본 연구의 결과를 통해 운동에 자주 참여하거나 운동에 참여하는 시간이 많은 학생일수록 자신의 신체에 대해서 긍정적으로 인식하고, 일상생활에서의 희망, 낙관성, 탄력성, 효능감을 가진다는 것을 확인하였으며, 초등학생들이 운동에 참여함으로써 향상된 신체적자기개념이 긍정심리자본의 향상으로 이어질 수 있다는 것을 실증적으로 규명하였다.
-------------	---

Epoch 수	Generation Sample
ver01_ep00	코로나19 이후 초등학생의 신체활동량과 신체적자기개념, 그리고 긍정심리자본의 관계를 탐색하고자 한 본 연구는 초등학교 6학년에 재학 중인 97명의 학생들을 대상으로 주간운동빈도, 주간운동시간, 신체적자기개념, 긍정심리자본을 측정하는 설문조사를 실시하였고, 수집된 자료를 바탕으로 기술통계분석, 상관분석, 구조방정식 모형 검증을 실시하였다. 본 연구를 수행한 결과, 초등학생들의 운동참여수준(주간운동빈도, 주간운동시간)은 신체적자기개념 및 긍정심리자본의 모든 하위요인들과 정적 상관관계가 나타났다. 또한 신체적자기개념은 운동참여수준과 긍정심리자본의 관계를 부분적으로 매개하는 것으로 나타났다. 본 연구의 결과를 통해 운동에 자주 참여하거나 운동에 참여하는 시간이 많은 학생일수록 자신의 신체에 대해서 긍정적으로 인식하고, 일상생활에서의 희망, 낙관성, 탄력성, 효능감을 가진다는 것을 확인하였으며, 초등학생들이 운동에 참여함으로써 향상된 신체적자기개념이 긍정심리자본의 향상으로 이어질 수 있다는 것을 실증적으로 규명하였다.
ver01_ep01	코로나19 이후 초등학생의 신체활동량과 신체적자기개념은 긍정심리자본의 모든 하위요인들과 정적 상관관계가 나타났다.
ver02_ep00	코로나19 이후 초등학생의 신체활동량과 신체적자기개념, 그리고 긍정심리자본의 관계를 탐색하고자 한 본 연구는 운동참여가 긍정심리자본에 영향력을 행사하는 과정에서 신체적자기개념의 매개효과를 검증하고자 하였다.
ver02_ep01	코로나19 이후 초등학생의 신체활동량과 신체적자기개념, 그리고 긍정심리자본의 관계를 탐색하고자 한 본 연구를 수행한 결과, 초등학생들의 운동참여수준(주간운동빈도, 주간운동시간)은 신체적자기개념 및 긍정심리자본의 모든 하위요인들과 정적 상관관계가 나타났고, 신체적자기개념은 운동참여수준과 긍정심리자본의 관계를 부분적으로 매개하는 것으로 나타났다.
ver02_ep02	코로나19 이후 초등학생의 운동참여수준은 신체적자기개념 및 긍정심리자본의 모든 하위요인들과 정적 상관관계가 나타났고 신체적자기개념은 운동참여수준과 긍정심리자본의 관계를 부분적으로 매개하는 것으로 나타났다.
ver03_ep00	코로나19 이후 초등학생의 신체활동량과 신체적자기개념, 긍정심리자본의 관계를 탐색하고자 한 본 연구는 운동참여가 긍정심리자본에 영향력을 행사하는 과정에서 신체적자기개념의 매개효과를 검증하고자 하였다.
ver03_ep01	코로나19 이후 초등학생의 신체활동량과 신체적자기개념, 그리고 긍정심리자본의 관계를 탐색하고자 운동참여가 긍정심리자본에 영향력을 행사하는 과정에서 신체적자기개념의 매개효과를 검증하고자 하였다.
ver03_ep02	코로나19 이후 초등학생의 운동참여수준은 신체적자기개념 및 긍정심리자본의 모든 하위요인들과 정적 상관관계가 나타났으며 운동참여가 긍정심리자본에 영향력을 행사하는 과정에서 신체적자기개념의 매개효과를 검증하고자 했다.
ver03_ep04	<b>코로나19</b> 이후 초등학생의 운동참여수준은 신체적자기개념 및 긍정심리자본의 모든 하위요인들과 정적 상관관계가 나타났고 신체적자기개념은 운동참여수준과 긍정심리자본의 관계를 부분적으로 매개하는 것으로 나타났다.

각 에포크 별로 주제문을 찾아 요약하고 있으나, 중심문장이 조금씩 다른 양상을 보임

에포크 증대 시 오탈자 발생

## IV. 프로젝트 수행 결과 – 최종모델 선정/Human Evaluation-candidates

Candidates	Generation Sample
label	2010년까지 9~10개의 최고품질 쌀 품종을 육성하는 것이 목표이며 소비자들로부터 최고품질이라는 평가를 받기 위해서는 우선 소비자의 만족도를 만족시켜야 한다.
ver02_ep02	농촌진흥청이 2005년 5월에 발표한 최고품질 벼 품종육성 5개년(2005~2009) 계획에서 삼광벼가 2003년도에, 고품 · 운광벼가 2004년도에 육성되었으나 현재 종자 증식단계에 있어 아직 농가에 보급되지 않고 있는 상황이어서 외국쌀과 경쟁이 불가하다고 판단하고 2010년까지 9~10개의 최고품질 쌀 품종을 육성하는 것을 목표로 하고 있다.
ver03_ep01	2000년부터 품종개발 목표가 고품질 쌀품종으로 바뀌었음에도 불구하고 2000년 이후 개발된 품종의 주요특성을 보면 밥맛이라는 특성을 가지고 있기는 하지만 수량성과 재배안정성이 라는 특성이 공통적으로 포함되어있다는 사실을 확인할 수 있다.
Candidates	Generation Sample
ver02_ep02	본 논문에서는 한국어 디비피디아의 웹 자원 접근성 개선과 한국어 워드넷 연결을 통한 활용도를 높이는 방안을 제시하며 한국어 디비피디아는 사람이 읽을 수 있는 상태의 리소스 페이지 제공이 가능하다.
ver03_ep01	본 논문에서는 한국어 디비디아의 웹 자원 접근성 개선과 한국어 워드넷 연결을 통한 활용도를 높이는 방안을 제시하며 사람이 읽을 수 있는 상태의 리소스 페이지 제공이 가능하다고 밝혔다.
ver02_ep02	전세계적으로 정보의 생산자와 유통경로가 다양해지면서 의도적으로 사실이 아닌 정보를 사실로 위장하여 전달하는 가 사회적 갈등 등의 이슈도 크게 부각되고 있다.
ver03_ep01	전세계적으로 정보의 생산자와 유통경로가 다양해지면서 의도적으로 사실이 아닌 정보를 사실로 위장하여 전달하는 가 적 오인과 사회적 갈등 등의 이슈도 크게 부각되고 있다.
ver02_ep02	지역 청년가들이 현장에서 느끼는 애로요인은 비즈니스 모델, 경영관리, 창업자금, 창업제도개선, 창업 인식개선 순으로 나타났다.
ver03_ep01	중앙정부는 일자리 창출을 위한 중요한 대안으로 청년 창업지원 정책을 확대하고 있지만 창업에 다양한 애로요인이 존재하여 본 연구에서는 지역 청년 창업가들이 현장에서 느끼는 실질적 인 애로요인이 무엇인지를 파악하고, 지역 청년창업의 애로 요인을 극복하기 위한 대안과 청년창업 활성화화를 위한 정책적 시사점을 도출하는데 목적이 있다.
ver02_ep02	코로나19로 인해 키오스크에서의 디지털 소외가 심화되고 있는데 이를 해결하기 위해 많은 연구가 이뤄지고 있지만 대부분이 인터페이스의 개선에 치중되어 있으며 고령자의 심리적 특징을 이해하고 사용자경험 과정에서 기술 수용을 돕기 위한 시스템적 디자인 방안을 제안하고자 한다
ver03_ep01	코로나19로 인해 키오스크에서의 디지털 소외가 심화되고 있는 가운데 본 연구에서는 고령자의 심리적 특징을 이해하고 사용자경험 과정에서 기술 수용을 돕기 위한 시스템적 디자인 방안을 제안하고자 한다.
ver02_ep02	기업가형 대학에 대한 관심이 높아지는 상황에서 대학생과 전임교원의 창업 성과를 중심으로 대학 차원의 영향 요인을 구명하는데 목적이 있다.
ver03_ep01	기업가형 대학에 대한 관심이 높아지는 상황에서 대학생과 전임교원의 창업 성과를 중심으로 대학 차원의 영향 요인을 구명하는데 목적이 있는 이 연구는 2015년부터 2018년까지 대학알리미 공시자료를 바탕으로 최소 2개년도 대학생 및 전임교원 창업자 수에 대한 자료가 확보되어 있는 대학을 대상으로 패널을 구축했다.

대부분 비슷하게 요약문을 생성했으나,  
전반적으로 Epoch 7(ver03\_ep01)의 품질이 일정하여,  
최종모델로 Epoch 7 모델 선정

## IV. 프로젝트 수행 결과 – 패키징 및 배포

패키지 구조 구축	<p>서비스 프로세스 구조화</p> <ul style="list-style-type: none"><li>- 사용자가 .txt 파일을 넣으면, 모델이 인식할 수 있도록 형태를 전처리</li><li>- 런버튼 실행으로 쉽게 결과를 확인할 수 있도록 모듈 설계<ul style="list-style-type: none"><li>- Summarize.Summarizer 모듈 : 탑재된 Fine-Tuned 모델을 불러와 자동으로 문서 요약</li><li>- Preprocess.Reader 모듈 : txt 파일의 형태를 모델이 인식할 수 있도록 형태를 한문장 형태로 전처리</li></ul></li><li>- Output으로 파일 내에서 직접 확인하거나, csv가 아닌 대부분의 유저에게 친숙한 Excel파일로 저장되도록 함</li></ul>
Product Manual 작성	<p>비전문가 대상으로 README 작성</p> <ul style="list-style-type: none"><li>- User friendly 하도록 모듈화하여 사용자의 가독성을 높임</li><li>- 모델 구조에 대해 이해하기 쉽도록 각 폴더의 역할을 기입</li><li>- 사용 주의사항을 기입하여 비전문가에게 발생할 수 있는 오류 방지</li></ul>
GitHub Repo 배포	<p>패키지 다운로드 혹은 git clone을 통해 패키지 사용할 수 있도록 배포</p> <p>대용량 모델 탑재</p> <ul style="list-style-type: none"><li>- Git-LFS를 통해 약 450MB의 fine-tuned model 탑재</li></ul>

Txt 파일을 넣으면 직접 확인 및 파일로 저장하는 논문/리포트요약모델 패키지(test.ipynb) 배포

## IV. 프로젝트 수행 결과 – 최종 서비스

### 서비스 파일

#### ▼ Paper/Report Summarization Service

- 본 서비스는 논문 및 레포트의 초록을 요약하여 주제문을 생성하는 서비스입니다.
- data 폴더에 txt 파일을 넣으면, 파일 내에서 혹은 result 폴더 내 excel 파일을 통해 결과를 확인하실 수 있습니다.
- [파일 불러오기 및 전처리]를 실행시킨 후 [요약문 생성 및 확인]을 실행해주시기 바랍니다.

#### ▶ 파일 불러오기 및 전처리

- 실행 시 파일을 불러와, 모델에 알맞은 형태로 자동으로 가공합니다
- 하단 [요약문 생성 및 확인]을 실행시키기 전 반드시 실행시켜주시기 바랍니다

[ ] ↓ 숨겨진 셀 2개

#### ▼ 요약문 생성 및 확인

- 파일이 한 개일 경우, 파일 내에서 바로 결과를 확인하실 수 있습니다
  - 여러 개일 경우 첫번째 파일의 결과만 출력됩니다
- 파일이 여러 개일 경우, result 폴더 내 excel 파일을 통해 결과를 확인하실 수 있습니다
  - 단, excel 파일이 열려있으면 제대로 동작하지 않습니다. 실행 전 excel 파일이 열려있지 않은지 확인해주세요.
  - 파일이 한 개인 경우에도 excel 파일을 통해 결과를 저장하실 수 있습니다

#### ▶ 결과 확인하기(파일이 한 개인 경우)

[ ] ↓ 숨겨진 셀 1개

#### ▶ 엑셀 파일로 저장하기(파일이 한 개 혹은 여러 개인 경우)

[ ] ↓ 숨겨진 셀 2개

### 요약 형태

A	B	C	D
	file_name	text	summary
0	sample0.tx	본 연구는	코로나19 이후 초등학생의 신체활동량과 신체적 자기개념, 그리고 긍정심리자본의 관계를 탐색하고자 운동참여가 긍정심리자본에 영향력을 행사하는 과정에서 신체적자기개념의 매개효과를 검증하고자 하였다.
1	sample1.tx	본 논문에서,	본 논문에서는 한국어 디비디아의 웹 자원 접근성 개선과 한국어 워드넷 연결을 통한 활용도를 높이는 방안을 제시하며 사람이 읽을 수 있는 상태의 리소스 페이지 제공이 가능하다고 밝혔다.
2	sample2.tx	창업의 중	창업의 중요성에 대한 인식이 높아지면서 인생의 진로를 창업으로 결정하는 예비 창업가들이 많아지고 있지만 창업 실패율은 여전히 높고, 창업성과가 미비하여 질적인 수준은 매우 낮다.

## V. 자체 평가 의견

의의	<p>완결된 형태의 요약문 생성</p> <ul style="list-style-type: none"><li>- 초록을 통해 요약문을 자동으로 생성하는 NLG모델을 구축하였음</li><li>- 추상적 요약모델을 사용하여 인간이 어색하지 않은 형태로 문장을 생성할 수 있었음</li><li>- 실제 문서에 어긋나지 않는 요약문을 생성하였음</li><li>- Fine Tune을 수행하고, 라이브러리를 활용해 사전학습모델과 그 결과를 비교함</li></ul> <p>평가지표를 자체적으로 구축</p> <ul style="list-style-type: none"><li>- 논문에 기반해 RDASS Score를 구축하여 일반적인 지표인 Rouge Score와 비교</li></ul> <p>패키징 및 모듈화를 통한 배포</p> <ul style="list-style-type: none"><li>- 사용자가 쉽게 사용할 수 있는 형태로 논문/리포트 요약 서비스를 제공할 수 있었음</li><li>- 주어진 데이터를 모델에 알맞게 변환하는 전처리 모듈, fine-tuned 모델을 자동으로 불러오는 요약 모듈 등 모듈화를 통해 사용자의 편의성과 가독성을 높였음</li></ul>
한계	<p>Fine Tune을 통한 성능 향상 미비</p> <ul style="list-style-type: none"><li>- 데이터셋에 논문/리포트 외에 여러 정보가 있어, text 및 label 형태가 일정하지 않아 fine-tuning이 사전학습모델에 큰 영향을 끼치지 않은 것으로 추측됨</li><li>- 이외에도 epoch 수 외 다양한 하이퍼파라미터를 튜닝할 필요가 있음</li><li>- Text 및 Label을 연구배경/방법/결과 세가지 중 하나로 분류해 각각 요약하는 전처리 과정을 추가해 요약문의 품질을 향상시킬 수 있을 것으로 추측됨</li></ul> <p>다양한 NLG 모델과의 비교</p> <ul style="list-style-type: none"><li>- KoBART 외에도 BART, BERT, KoBERT 모델 등 다양한 모델과의 성능 비교 필요</li></ul>
보완점	<ul style="list-style-type: none"><li>- 데이터 정량적 확보 및 품질 향상</li><li>- 연구배경/연구방법/연구결과로 나누어 요약하는 전처리 파이프라인 추가</li><li>- Fine-Tune 방법 다양화 : beam size, batch size 등</li><li>- 다양한 NLG 모델 및 다른 SOTA 모델(뉴스요약 - Pegasus+Xsum) 과의 결과물 비교</li></ul>

